1977-03

A note on a confidence interval for an interclass mean

Jayachandran, T.

Wiley Online

Jayachandran, T. "A note on a confidence interval for an interclass mean." Naval Research Logistics Quarterly 24.1 (1977): 197-200.

http://hdl.handle.net/10945/60197

A NOTE ON A CONFIDENCE INTERVAL FOR AN INTERCLASS MEAN

T. Jayachandran*

Naval Postgraduate School
Monterey, California

ABSTRACT

An exact confidence interval for an interclass mean, that is, the mean of a composite sample made up several subsamples of unequal sizes n_i is presented.

1. INTRODUCTION

Suppose $X_{ij}, j=1, 2, \ldots, n_i; i=1, 2, \ldots, k$ is a composite sample of size

$$N=\sum_{i=1}^{k} n_i$$

that is comprised of k subsamples of sizes n_i. The i^{th} subsample is a random sample from a normal distribution $N(a_i,\sigma^2)$ and the a_i's, $i=1, 2, \ldots, k$ are assumed to be independent and identically distributed $(i \cdot i \cdot d)$ as $N(\mu, \sigma_a^2)$. μ is sometimes known as the interclass mean and the problem considered in this note is the construction of a confidence interval for μ. Some practical uses of such an interval are discussed in a paper by Long [1]. For example, the composite sample may be measurements on a characteristic of the output of a factory made on different days. μ would correspond to the true mean value of the characteristic being studied.

The construction of a confidence interval for μ is straightforward if all the n_i are equal. For unequal subsample sizes Long [1] obtained an approximate interval that he shows to be reasonably accurate. The procedure proposed in this paper leads to an exact confidence interval for μ in the case of unequal n_i.

2. PRELIMINARIES

The observations $X_{ij}, j=1, 2, \ldots, n_i; i=1, 2, \ldots, k$ may be assumed to satisfy the variance components model

(2.1) $$Y_{ij}=a_i+e_{ij},\ j=1, 2, \ldots, n;\ i=1, 2, \ldots, k$$

*This research was supported by the Office of Naval Research as part of the Foundation Research Program at the Naval Postgraduate School.

where the a_i's are $i{\cdot}i{\cdot}d.$ $N(\mu\sigma_a^2)$, the e_{ij}'s are $i{\cdot}i{\cdot}d.$ $N(0, \sigma^2)$ and the a_i's and e_{ij}'s are mutually independent. Let

$$\overline{X}_i=\sum_{j=1}^{n_i}\frac{X_{ij}}{n_i} \text{ and } \overline{X}=\sum_{i=1}^{k}\frac{\overline{X}_i}{k}$$

and

$$S^2=\sum_{i=1}^{k}\frac{(\overline{X}_i-\overline{X})^2}{k-1}$$

Then, $\overline{X}_i$, $i=1, 2, \ldots, k$ are independent and normally distributed

$$N\left(\mu, \sigma_a^2+\frac{\sigma^2}{n_i}\right).$$

If $n_i=n$ for all i the $\overline{X}_i$ will constitute a random sample from a normal distribution

$$N\left(\mu, \sigma_a^2+\frac{\sigma^2}{n}\right).$$

Thus,

$$T=\frac{k^{1/2}(\overline{X}-\mu)}{S} \tag{2.2}$$

has a student's t distribution with (k-1) degrees of freedom, and this leads to a confidence interval or μ.

For the case where the n_i are not all equal, Long [1] obtained an approximate confidence interval for μ by treating T as a student's t variable. On the basis of a study of the exact distribution fof T for $k=2$, 3 and an examination of the moments of T for large values of k, Long [1] concludes that the t approximation is fairly reliable. He also points out that the approximation is likely to go wrong if there are wide variations in the subsample sizes n_i or if σ^2 is large relative to σ_a^2.

3. EXACT CONFIDENCE INTERVAL WITH UNEQUAL SUBSAMPLE SIZES

For unequal subsample sizes, the $\overline{X}_i$ are not identically distributed since the variance of $\overline{X}_i$ is

$$\sigma_a^2+\frac{\sigma^2}{n_i},$$

even though all of them have the same mean μ. Thus, the construction of a student's t variable, based on $\overline{X}_i$ only, independent of the nuisance parameters $\sigma_a^2+\sigma^2/n_i$, is not possible. To get around this difficulty let

$$Z_i=C_{i1}X_{i1}+C_{i2}\overline{X}_i \qquad i=1, 2, \ldots k \tag{3.1}$$

where

$$C_{i1}=\left(\frac{n_iC-1}{n_i-1}\right)^{1/2}$$

$$C_{i2}=1-C_{i1}$$

and

$$C = \frac{1}{\min n_i}.$$

Then, as shown in the appendix,

$$E(Z_i) = \mu$$

$$V(Z_i) = \sigma_a^2 + C\sigma^2 = \sigma_a^2 + \frac{\sigma^2}{\min n_i}$$

If

$$\bar{Z} = \frac{1}{k} \sum_{i=1}^{k} Z_i$$

and

$$S_Z^2 = \frac{1}{k_{-1}} \sum_{i=1}^{k} (Z_i - \bar{Z})^2$$

then

$$T = \frac{k^{1/2}(\bar{Z} - \mu)}{S_Z} \tag{3.2}$$

has a student's t-distribution with k–1 degrees of freedom. An exact confidence interval for μ is now obtainable in the usual way.

4. DISCUSSION

The exact procedure proposed in Section 3 is somewhat ad hoc in nature in the sense that the first observation X_{i1} in each subsample plays a prominent role. To avoid any systematic bias that may creep in, it would be preferable to randomly permute the observations in each of the k subsamples before applying the procedure. If all the subsample sizes n_i are equal to n, the statistic (3.2) reduces to the usual T statistic (2.2).

5. APPENDIX

Under the assumptions of the variance components model (2.1)

$$E(X_{ij}) = \mu \qquad j = 1, 2, \ldots, n_i;\ i = 1, 2, \ldots, k$$

$$\text{Var}\ (X_{ij}) = \sigma^2 + \sigma_a^2$$

and

$$\text{Cov}\ (X_{ij}, X_{i'j'}) = \begin{cases} 0 & i \neq i' \\ \sigma_a^2 & i = i',\ j \neq j' \end{cases}$$

It follows

$$E(\bar{X}_i) = \mu$$

$$V(\bar{X}_i) = \sigma_a^2 + \frac{\sigma^2}{n_i} \qquad i = 1, 2, \ldots, k$$

and

$$\text{Cov}\ (X_{ij}, \bar{X}_i) = \sigma_a^2 + \frac{\sigma^2}{n_i} \qquad j = 1, 2, \ldots, n_i$$

Therefore, if

$$Z_t = C_{t1} X_{t1} + C_{t2} \overline{X}_t$$

then

$$E(Z_t) = (C_{t1} + C_{t2})\mu = \mu$$

and

$$V(Z_t) = C_{t1}^2 V(X_{t1}) + C_{t2}^2 V(\overline{X}_t) + 2C_{t1}C_{t2} \operatorname{Cov}(X_{t1}, \overline{X}_t)$$

$$= C_{t1}^2(\sigma^2 + \sigma_a^2) + (C_t^2 + 2C_{t1}C_{t2})\left(\sigma_a^2 + \frac{\sigma^2}{n_t}\right).$$

substitution of

$$C_{t1} = \left(\frac{n_t C - 1}{n_t - 1}\right)^{1/2},\ C_{t2} = 1 - C_{t1} \text{ and } C = \frac{1}{\min n_t}$$

results in

$$E(Z_t) = \mu \text{ and } V(Z_t) = \sigma_a^2 + \frac{\sigma^2}{\min n_t}.$$

REFERENCES

1. Long, W. M., "Estimation Problems When a Simple Type of Heterogeneity is Present in the Sample," Biometrika, *38*, 90–101 (1951).

*U.S. GOVERNMENT PRINTING OFFICE: 1977240-830/1 1-3